What is claimed is:

1. A method for extracting infinite ambiguity from an input finite-state transducer (FST) having a plurality of states and a plurality of arcs, comprising the steps of:

assigning to each state a set of epsilon loops and a unique diacritic representative of the set; each epsilon loop in the set of epsilon loops beginning and ending at a corresponding state; said assigning step defining a first representation of the input FST and a second representation of the input FST;

building a first factor by inserting into the first representation of the input FST one auxiliary state for each state with a non-empty set of epsilon loops; wherein each auxiliary state has an arc that leads from the auxiliary state to the corresponding state and emits the corresponding unique diacritic when traversed;

removing from the first factor at least one epsilon loop without removing the arcs corresponding to the epsilon loop;

building a second factor by inserting into the second representation of the input FST two auxiliary arcs for each state with a non-empty set of epsilon loops; wherein the two auxiliary arcs are labeled with a diacritic, and wherein a first of the auxiliary arcs leads from an initial state to its corresponding state, and a second of the auxiliary arcs leads from its corresponding state to a final state; and

removing from the second factor all paths having partial epsilon loops; and

mapping each diacritic in the second factor to a corresponding set of epsilon loops.

2. The method of claim 1, further comprising the steps of:

minimizing the first factor; and

minimizing the second factor.

3. The method of claim 2, further comprising the steps of:

concatenating at least one boundary symbol to the input FST; and

minimizing the input FST.

4. The method of claim 2, wherein said step of removing at least one epsilon loop from the first factor without removing arcs corresponding to the epsilon loop further comprises the steps of:

temporarily replace each arc of the epsilon loop with a diacritic to define a sequence of diacritics;

formulate a constraint that disallows the sequence of diacritics;

compose the constraint with the first factor; and

replace any remaining diacritics with an epsilon symbol.

5. The method of claim 4, wherein said step of removing all paths having partial epsilon loops from the second factor further comprises the step of mapping any sequence of two identical diacritics to itself and inserting a corresponding epsilon loop there between.

6. The method of claim 1, wherein partial epsilon loops are paths with an input side having one of the unique diacritics occurring only once.

7. The method of claim 1, wherein said step of removing from the second factor all paths having epsilon loops further comprises removing arcs from the second factor.

8. The method of claim 1, wherein said step of removing from the second factor all paths having epsilon loops further comprises rearranging arcs in the second factor.

9. The method of claim 1, wherein a selected state has a non-empty set of epsilon loops if starting at the selected state a sequence of arcs, each having an epsilon label, is traversed in the input FST that terminates at the selected state.

10. The method of claim 1, further comprising the step of factoring the first factor into a functional FST and a fail-safe FST.

11. The method of claim 10, wherein the functional FST, the fail-safe FST, and the second factor are adapted for performing language processing.

12. The method of claim 11, wherein the language processing comprises one of tokenization, phonological analysis, morphological analysis, disambiguation, spelling correction, and shallow parsing.

13. The method of claim 10, wherein the functional FST, the fail-safe FST, and the second factor form part of a lexical transducer.

14. An apparatus for extracting infinite ambiguity from an input finite-state transducer (FST) having a plurality of states and a plurality of arcs, comprising:

means for assigning to each state a set of epsilon loops and a unique diacritic representative of the set; each epsilon loop in the set of epsilon loops beginning and ending at a corresponding state; said assigning means defining a first representation of the input FST and a second representation of the input FST;

means for building a first factor by inserting into the first representation of the input FST one auxiliary state for each state with a non-empty set of epsilon loops; wherein each auxiliary state has an arc that leads from the auxiliary state to the corresponding state and emits the corresponding unique diacritic when traversed;

means for removing from the first factor at least one epsilon loop without removing the arcs corresponding to the epsilon loop;

means for building a second factor by inserting into the second representation of the input FST two auxiliary arcs for each state with a non-empty set of epsilon loops; wherein the two auxiliary arcs are labeled with a diacritic, and wherein a first of the auxiliary arcs leads from an initial state to its corresponding state, and a second of the auxiliary arcs leads from its corresponding state to a final state; and

means for removing from the second factor all paths having partial epsilon loops; and

means for mapping each diacritic in the second factor to a corresponding set of epsilon loops.

15. The apparatus of claim 14, further comprising means for factoring the first factor into a functional FST and a fail-safe FST.

16. The apparatus of claim 15, wherein the functional FST, the fail-safe FST, and the second factor are adapted for performing language processing.

17. The apparatus of claim 16, wherein the language processing comprises one of tokenization, phonological analysis, morphological analysis, disambiguation, spelling correction, and shallow parsing.

18. The apparatus of claim 15, wherein the functional FST, the fail-safe FST, and the second factor form part of a lexical transducer.